第47卷 第4期
2024年4月

干旱区地理
ARID LAND GEOGRAPHY

Vol. 47 No. 4
Apr. 2024

基于Sentinel-2时序数据的新疆焉耆盆地农作物遥感识别与评估

张旭辉[1]， 玉素甫江·如素力[1,2]， 仇忠丽[1]
亚夏尔·艾斯克尔[1]， 阿卜杜热合曼·吾斯曼[1]

（1. 新疆师范大学地理科学与旅游学院流域信息集成与生态安全实验室，新疆 乌鲁木齐 830054；
2. 新疆干旱区湖泊环境与资源重点实验室，新疆 乌鲁木齐 830054）

摘 要：为及时准确地获取干旱区农作物种植信息，研究借助PIE-Engine Studio平台，以新疆焉耆盆地为研究区，基于2022年Sentinel-2影像和1948个野外定位采样数据提取农作物生育期内14种植被指数，使用See5.0决策树、随机森林（Random forest，RF）和多元回归（Multiple regression，MR）模型优选特征参数，结合支持向量机（Support vector machine，SVM）算法构建5种分类模型和5种样方分割方案进行农作物种植信息提取，通过目视解译和混淆矩阵对比分析分类结果，确定最佳分类方案。结果表明：（1）所有分类模型的总体精度（OA）和Kappa系数均在92.20%和0.9037以上，说明在PIE平台中使用SVM算法提取农作物信息是可行的。（2）SVM-有红边的OA和Kappa系数均值为93.77%和0.9236，比SVM-无红边方法提高了0.96%和0.0120。（3）相比于SVM-有红边方法，植被指数的引入提高了SVM-RF、SVM-MR和SVM-See5.0的OA和Kappa系数。（4）5种分类模型的OA和Kappa系数均值的大小关系为：SVM-RF>SVM-MR>SVM-See5.0>SVM-有红边>SVM-无红边，表明红边波段和植被指数的加入显著提高了农作物识别的精度，其中SVM-RF（8:2）为最佳分类模型，OA和Kappa系数分别为98.72%和0.9866。研究结果可为准确快速获取大尺度干旱区农作物信息提供新的思路和参考依据。

关 键 词：农作物；Sentinel-2；支持向量机；PIE-Engine Studio；焉耆盆地
文章编号：1000－6060(2024)04－0672－12(0672～0683)

农作物种植信息是进行农作物产业结构调整和优化的重要基础和前提[1]。焉耆盆地是新疆重要的特色农产品生产基地，为国家供应大量粮食，但粮食生产受到荒漠化、土壤盐碱化及干旱缺水等胁迫，作物种植结构复杂、生态环境脆弱[2]。因此，及时准确地掌握焉耆盆地农作物种植信息对促进干旱区农业可持续发展和保障国家粮食安全具有重要意义[3-4]。

遥感技术因其覆盖面大、探测速度快及成本低的特点，为及时准确地获取大面积农作物种植信息提供了新的技术手段[5-6]。目前，基于单一影像源提取农作物种植结构信息的方法操作简单，但不易捕捉作物种植“最佳识别期”影像[7]；基于多时序影像源提取农作物种植信息的方法充分利用了作物季相节律特征，成为当下农作物种植信息提取的主流方法[3]；在提取农作物信息时，多特征参数法更适用于复杂的农作物种植区，基于多特征参数的统计模型法在一定程度上解决了混合像元问题[8]；遥感影像与统计数据融合法可获得大尺度农作物结构种植图，但因较低的制图分辨率使得产品区域适宜性

收稿日期：2023-06-06； **修订日期：**2023-07-24
基金项目：国家自然科学基金项目（U1703341，41764003）；自治区科技创新基地建设计划项目（2020D04039）资助
作者简介：张旭辉（1997-），男，硕士研究生，主要从事空间信息分析与应用研究. E-mail: philamour@163.com
通讯作者：玉素甫江·如素力（1975-），男，博士，教授，主要从事流域水文与生态环境遥感研究. E-mail: Yusupjan@xjnu.edu.cn

较差[9-10]。在机器学习方法中支持向量机算法泛化能力强，且有较好的鲁棒性，在遥感信息提取上被广泛应用[11-12]。梁继等[6]、贾银江等[13]和梁习卉子等[14]使用SVM算法及其改进后的分类器对区域农作物信息进行获取，取得较好的分类效果（分类精度均高于90%）。边增淦等[15]、田鑫等[16]和郭其乐等[17]研究指出，使用机器学习方法进行遥感信息提取时加入光谱特征、纹理特征及物候特征可以提高农作物分类的精度。随着遥感大数据时代的到来，遥感云平台为快速处理和分析海量遥感数据提供了机遇[18]。刘通等[19]、潘力等[20]和姜伊兰等[21]借助GEE平台分别对辽宁盘锦市、淮河流域和开封市杞县的农作物种植信息进行提取，研究结果为当地农作物信息提取和监测提供了参考依据。可是当下GEE平台国内用户不便正常使用，且使用过程和处理结果的下载，受网络条件的影响很大[22]。国产遥感云平台PIE-Engine Studio为共享数据、算法和计算资源构建了一个开放的环境，以实现对遥感数据的按需访问和海量数据的快速处理[22]。然而目前使用PIE-Engine Studio云计算平台结合机器学习方法应用于干旱区农作物识别的研究还相对较少，对于该平台在干旱区农作物识别的适用性还待进一步研究，且基于该平台的时间序列Sentinel-2农作物识别的研究还鲜有涉及，Sentinel-2数据的高时间分辨率在农作物分类识别中的优势还需进一步明确。

文章以新疆焉耆盆地为研究区，利用PIE-Engine Studio平台，基于Sentinel-2时序数据和野外定位采样数据，使用支持向量机算法对2022年焉耆盆地的农作物种植信息进行提取和精度评估，以期获取最佳的农作物信息提取方案，为干旱区农作物种植信息的快速、精确提取提供新的思路和参考依据。

1　材料与方法

1.1　研究区概况

焉耆盆地位于新疆巴音郭楞蒙古自治州境内（85°45′25″~87°26′11″E，41°40′11″~42°25′46″N），塔里木盆地的东北侧，面积约为7500 km²，地势西北高东南低，地理区域包含焉耆回族自治县、和静县、和硕县、博湖县及新疆生产建设兵团部分团场（图1）。该地区四季分明，光照时间充足，昼夜温差大，水热资源丰富，属典型的干旱区绿洲气候。农作物类型有番茄、辣椒、玉米、小麦、甜菜和棉花等，种植结构较为复杂，其中研究区盛产番茄和辣椒，是新疆重要的特色农产品生产基地。

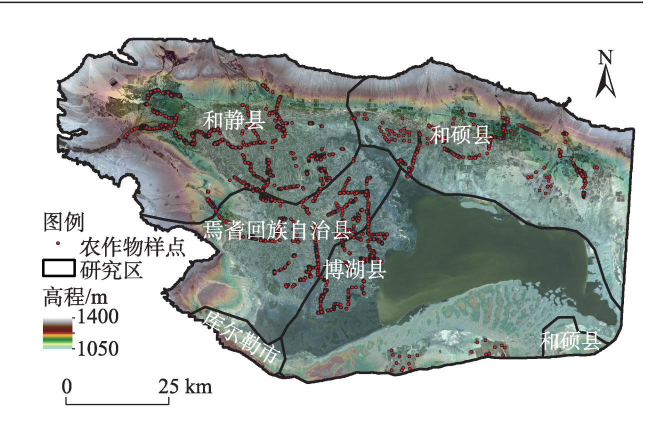


图1　研究区地理位置及作物样本点分布

Fig. 1　Geographical location of the study area and distribution of crop sample points

1.2　数据来源

1.2.1　Sentinel-2数据集　本文使用的Level-2A级Sentinel-2多光谱影像数据集源自欧洲航天局数据中心，该数据由Sentinel-2A和Sentinel-2B组成，具有空间分辨率高、重访周期短、光谱波段通道数目多以及波段宽度窄等特点[23]。为方便研究不同农作物生育期影像变化，研究在PIE平台中选择2022年3—11月的遥感影像，云量覆盖百分比在10%以下，共9幅（表1）。

1.2.2　样本数据　研究团队于2022年7月26日至2022年8月4日前往焉耆盆地对农作物种植情况进行调查，采用拍照和GPS定位的方式，获取1948个典型地物样点，其中农作物包括5种经济作物（甘草、葵花、芦苇、棉花和葡萄）、3种粮食作物（水稻、小麦和玉米）、3种蔬菜（番茄、辣椒、甜菜）和果树作物（果林）。在采样过程中使用GPS定位作物样点的经纬度和海拔，并且对地物的周边生长环境及灌溉方式进行记录，方便观察不同农作物的分布特征和生长环境。在分类过程中将样点随机分割成训练样本和检验样本，得到5种样点分割方案，即5:5、6:4、7:3、8:2和9:1（表2）。

1.3　研究方法

1.3.1　分类特征构建

（1）Sentinel-2光谱波段

Sentinel-2数据共有13个波段，研究选择其中用于土地类型检测、陆表检测、植被监测和环境检

表1 研究区Sentinel-2影像数据

Tab. 1 Sentinel-2 image data for the study area

序号	传感器类型及产品序号	影像获取时间（年-月-日）	光谱波段	空间分辨率/m	云量/%
1	S2_L2A_45TUG_20220304	2022-03-04	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
2	S2_L2A_45TUG_20220403	2022-04-03	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
3	S2_L2A_45TUG_20220503	2022-05-03	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
4	S2_L2A_45TUG_20220605	2022-06-05	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
5	S2_L2A_45TUG_20220705	2022-07-05	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
6	S2_L2A_45TUG_20220804	2022-08-04	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
7	S2_L2A_45TUG_20220903	2022-09-03	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
8	S2_L2A_45TUG_20221003	2022-10-03	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10
9	S2_L2A_45TUG_20221102	2022-11-02	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10、20	<10

表2 研究区地物样本统计表

Tab. 2 Statistics of feature samples in the study area

样本类别	样本总数	5:5		6:4		7:3		8:2		9:1	
		训练数	检验数	训练数	检验数	训练数	检验数	训练数	检验数	训练数	检验数
番茄	190	95	95	114	76	133	57	152	38	171	19
甘草	70	35	35	42	28	49	21	56	14	63	7
葵花	180	90	90	108	72	126	54	144	36	162	18
辣椒	289	145	144	173	116	202	87	231	58	260	29
芦苇	120	60	60	72	48	84	36	96	24	108	12
棉花	125	63	62	75	50	88	37	100	25	113	12
葡萄	190	95	95	114	76	133	57	152	38	171	19
果林	70	35	35	42	28	49	21	56	14	63	7
水稻	180	90	90	108	72	126	54	144	36	162	18
甜菜	289	145	144	173	116	202	87	231	58	260	29
小麦	120	60	60	72	48	84	36	96	24	108	12
玉米	125	63	62	75	50	88	37	100	25	113	12

测的波段，共10个，分别是B2、B3、B4、B5、B6、B7、B8、B8a、B11和B12[23]。

（2）植被指数构建

植被指数因其经过比值处理，可以部分消除因太阳高度角、卫星观测角、地形、大气程辐射等所带来的影响，被广泛用于农作物分类与识别、长势监测与产量估算等方面。研究利用Sentinel-2的波段构建多种植被指数（表3）。

1.3.2 *特征参数优选方法* 构建较多的特征参数在一定程度上提高农作物识别精度，但过多的特征参数往往会出现信息冗余现象，影响分类模型的效率和精度[17]。研究采用以下特征参数优选方法：（1）See5.0决策树算法根据提供最大信息增益的字段对样本数据进行分割，并对决策树的叶子进行裁剪或合并以提高分类精度，最后确定每个叶子的最佳阈值，从而得到每个特征参数的重要性得分[24]。（2）随机森林（Random forest，RF）方法通过对对象和变量进行采样来构建预测模型，即生成多个决策树并依次对对象进行分类，最后将每个决策树的分类结果汇总，所有预测类别的复数就是随机森林预测的对象类别，分类精度得到提高，同时也计算了参与分类的不同特征的重要性[25]。（3）多元回归（Multiple regression，MR）模型指含有多个解释变量的线性回归模型，用于解释被解释的变量与其他多个变量解释变量之间的线性关系[26]。

1.3.3 *分类方法与分类方案* 支持向量机（Support vector machines，SVM）算法是基于统计学习的机器学习方法，该算法利用结构风险最小化原理，在确

表3 植被指数计算公式

Tab. 3 Vegetation index calculation formula

序号	植被指数	计算公式
1	NDVI	(B8 − B4)/(B8 + B4)
2	$NDVI_{705}$	(B5 − B4)/(B5 + B4)
3	$NDVI_{740}$	(B6 − B4)/(B6 + B4)
4	$NDVI_{783}$	(B7 − B4)/(B7 + B4)
5	GI	B3/B4
6	VIgreen	(B3 − B4)/(B3 + B4)
7	GNDVI	(B8 − B3)/(B8 + B3)
8	RDVI	$(B8-B4)/\sqrt{B8+B4}$
9	TVI	0.5[120(B8 − B3) − 200(B4 − B3)]
10	EVI	$G(\frac{B8-B4}{B8+C_1\times B4-C_2\times B2+L})$
11	SRI	B8/B4
12	MSR	$[(B8/B4)-1]/[(B8/B4)^{0.5}+1]$
13	MCARI	[(B8 − B4) − 0.2(B8 − B3)]×(B8/B4)
14	TCARI	3[(B8 − B4) − 0.2(B8 − B3)×(B8/B4)]

注：NDVI、$NDVI_{705}$、$NDVI_{740}$、$NDVI_{783}$、GI、VIgreen、GNDVI、RDVI、TVI、EVI、SRI、MSR、MCARI和TCARI分别代表归一化植被指数、红边705归一化植被指数、红边740归一化植被指数、红边783归一化植被指数、绿度指数、绿色植被指数、绿色归一化植被指数、重正化差异植被指数、三角植被指数、增强植被指数、简单比率指数、修正简单比率、改良叶绿素吸收反射率指数和转换叶绿素吸收比指数；B2、B3、B4、B5、B6、B7和B8分别代表Sentinel-2影像的蓝光波段、绿光波段、红光波段、红边705波段、红边740波段、红边783波段和近红外波段。G为增益因子；C_1为第一大气修正系数，用于校正大气中的气溶胶影响；C_2为第二大气修正系数，用于校正大气中的水汽影响；L为土壤调节因子，用于校正土壤背景的影响。下同。

保误差最小的情况下，通过降低模型泛化误差的上限来提高分类模型的泛化能力[27]，SVM算法的径向基核函数可将训练样本映射到高维空间，从而便于进行多特征参数的农作物分类。本文基于Sentinel-2时序数据构建特征参数，借助R语言优选特征参数，最后使用SVM算法构建5种分类方案进行农作物信息提取（表4）。

1.3.4 *精度评价* 使用目视解译和混淆矩阵对分类结果进行精度评价，采用以下评价指标评估模型的精度，总体精度（Overall accuracy，OA）、Kappa系数、用户精度（User's accuracy，UA）、制图精度（Producer's accuracy，PA）、F1分数（F1 score）、错分误差（Commission error，CE）和漏分误差（Omission error，OE）。计算公式如下：

$$OA=\frac{\sum_{i=1}^{n}P_{ii}}{N}\times 100\% \tag{1}$$

$$Kappa=\frac{N\sum_{i=1}^{n}P_{ii}-\sum_{i=1}^{n}(P_{i+}\times P_{+i})}{N^2-\sum_{i=1}^{n}(P_{i+}\times P_{+i})} \tag{2}$$

$$UA=\frac{P_{ii}}{P_{i+}}\times 100\%,\ PA=\frac{P_{ii}}{P_{+i}}\times 100\% \tag{3}$$

$$F1\ score=\frac{2PA\times UA}{PA+UA}\times 100\% \tag{4}$$

$$CE=\frac{P_{i+}-P_{ii}}{P_{i+}}\times 100\% \tag{5}$$

$$OE=\frac{P_{+i}-P_{ii}}{P_{+i}}\times 100\% \tag{6}$$

式中：n为类别数量；N为验证样本总数；P_{ii}为每类中正确分类样本的数量，是i行i列上的值；P_{i+}为分类器将验证样本分为某一类别的总数；P_{+i}为某一类别验证样本总数。

2 结果与分析

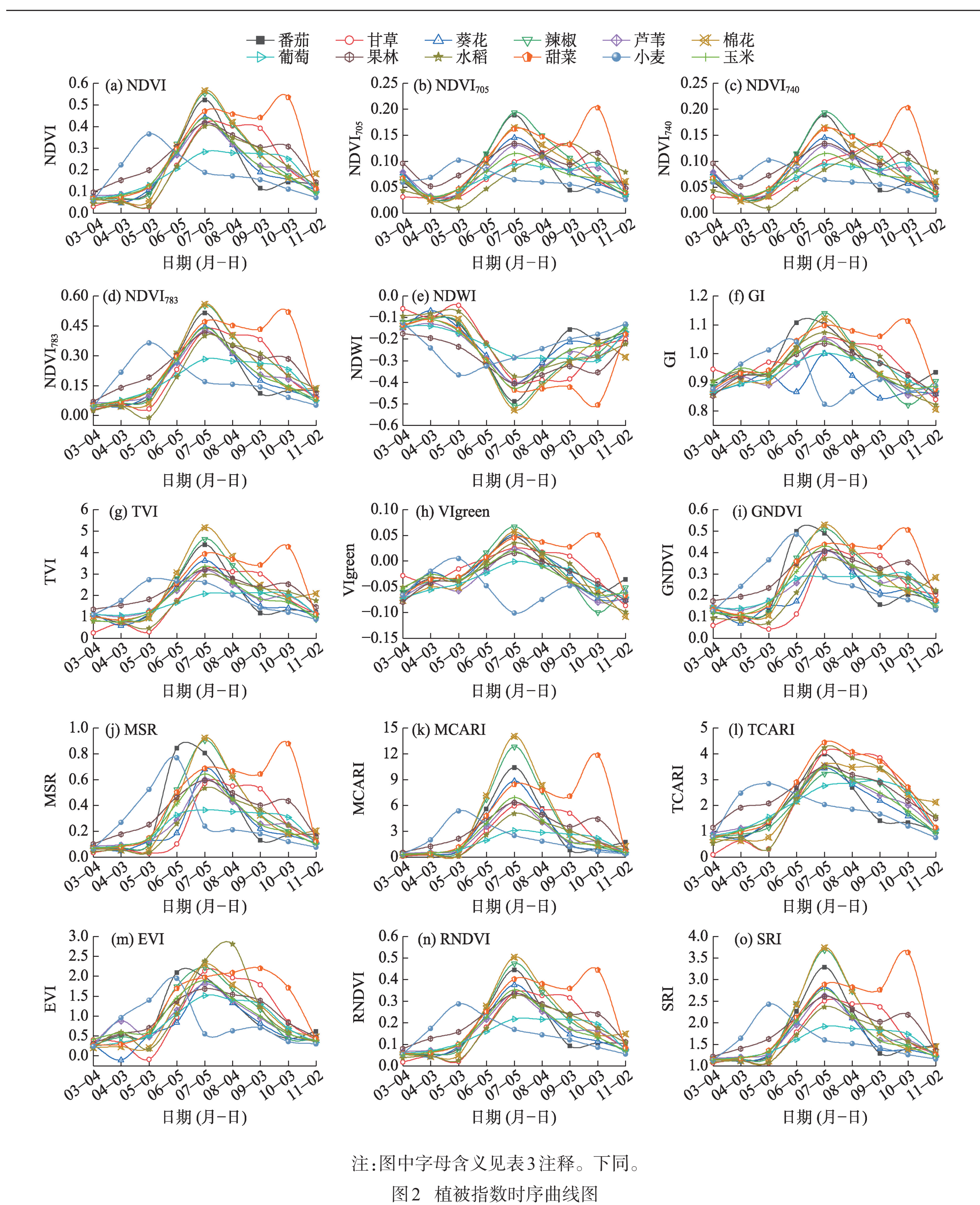
2.1 农作物植被指数时序特征

基于Sentinel-2时序数据计算植被指数，得到研究区农作物植被指数时序变化曲线（图2）。由图2可知，不同农作物因各自生物学差异在各植被指数中呈现不同的曲线波动趋势。由图2a可知，研究区农作物NDVI值总体呈先上升后下降的变化趋

表4 农作物分类方案

Tab. 4 Crop classification schemes

分类模型	特征参数	波段数目	样本分割方案
SVM-无红边	B2、B3、B4、B8、B8a、B11、B12	7	5:5、6:4、7:3、8:2、9:1
SVM-有红边	B2、B3、B4、B5、B6、B7、B8、B8a、B11、B12	10	5:5、6:4、7:3、8:2、9:1
SVM-See5.0	B8、EVI、B5、B6、TVI、B7、TCARI、$NDVI_{705}$、GI、B3、B2、VIgreen、B4、$NDVI_{740}$、NDWI、RNDVI、NDVI、$NDVI_{783}$、MCARI、B11、B8a、SRI	22	5:5、6:4、7:3、8:2、9:1
SVM-RF	B8、$NDVI_{705}$、B3、B6、TVI、B11、B2、B4、EVI、B7、MCARI、RNDVI、B8a、VIgreen、GI、B5、$NDVI_{740}$、TCARI、GNDVI、NDWI	20	5:5、6:4、7:3、8:2、9:1
SVM-MR	MCARI、NDVI、TCARI、SRI、TVI、VIgreen、$NDVI_{783}$、MSR、B5、B4、B8a、EVI	12	5:5、6:4、7:3、8:2、9:1



注:图中字母含义见表3注释。下同。

图2 植被指数时序曲线图

Fig. 2 Time-serie graphs of vegetation index

势,其中小麦和甜菜的NDVI值与其他农作物差异显著,容易区分;引入红边波段后,各农作物的NDVI值在7、8、9月差异较为明显(图2b~d),其中小麦的NDVI值在5月达到峰值,与其他农作物差异明显,容易区分;甜菜的NDVI值在10月达到峰值,与其他农作物相差较大,容易区分;与NDVI值相比,不同农作物的$NDVI_{705}$、$NDVI_{740}$和$NDVI_{783}$曲线分离较好,不同农作物的物候特征更直观、更清晰,这是因为红边是绿色植被反射光谱在680~780 nm之间的最大斜率点,因此更能反映农作物的微小变化。

由图2m可知，水稻的EVI曲线在8月达到峰值，与其他农作物相比差异明显；由图2n~o可知，棉花的RNDVI和SRI曲线在7月达到峰值，与其他农作物相比差异明显。除NDWI（图2e）外，其他农作物植被指数（图2f~o）呈现先增加后减少趋势，由于植被指数与农作物本身的特性有关，不同农作物的物候差异导致植被指数在增加、减少和峰值出现的时间等方面有差异。

2.2　特征参数重要性评价

研究利用See5.0、RF和MR方法得到特征参数的重要性评价结果（图3）。由图3可知，3种方法得到的特征参数重要性排序均不相同。See5.0和RF方法的结果中Sentinel-2近红外波段（B8）的重要性得分最高，说明近红外波段对农作物分类结果的贡献最大，对分类精度影响颇深。对比3种重要性结果发现，排名前10的特征参数中均有NDVI或红边NDVI，说明NDVI或者红边NDVI可以很好地反映出不同农作物的生长状态和覆盖度，对农作物分类结果贡献较大。因此根据特征参数重要性评价结果选取3个特征参数集：

（1）See5.0算法的特征重要性得分和累计百分比（图3a），获取重要性得分大于0的特征参数，符合条件的有B8、EVI、B5、B6、TVI、B7、TCARI、$NDVI_{705}$、GI、B3、B2、VIgreen、B4、$NDVI_{740}$、NDWI、RNDVI、NDVI、$NDVI_{783}$、MCARI、B11、B8a和SRI，共22个。

（2）RF的特征参数重要性评价（图3b），选取累计百分比大于90%的特征参数，符合条件的有B8、$NDVI_{705}$、B3、B6、TVI、B11、B2、B4、EVI、B7、MCARI、RNDVI、B8a、VIgreen、GI、B5、$NDVI_{740}$、TCARI、GNDVI和NDWI，共20个。

（3）MR模型的特征参数重要性评价（图3c），选取回归系数大于0的特征参数，符合条件的有MCARI、NDVI、TCARI、SRI、TVI、VIgreen、$NDVI_{783}$、MSR、B5、B4、B8a和EVI，共12个。

2.3　精度评价

研究对25种分类结果进行精度验证与对比，结果见图4：（1）所有农作物分类模型的OA均大于92.20%，Kappa系数均大于0.9037，说明在PIE平台

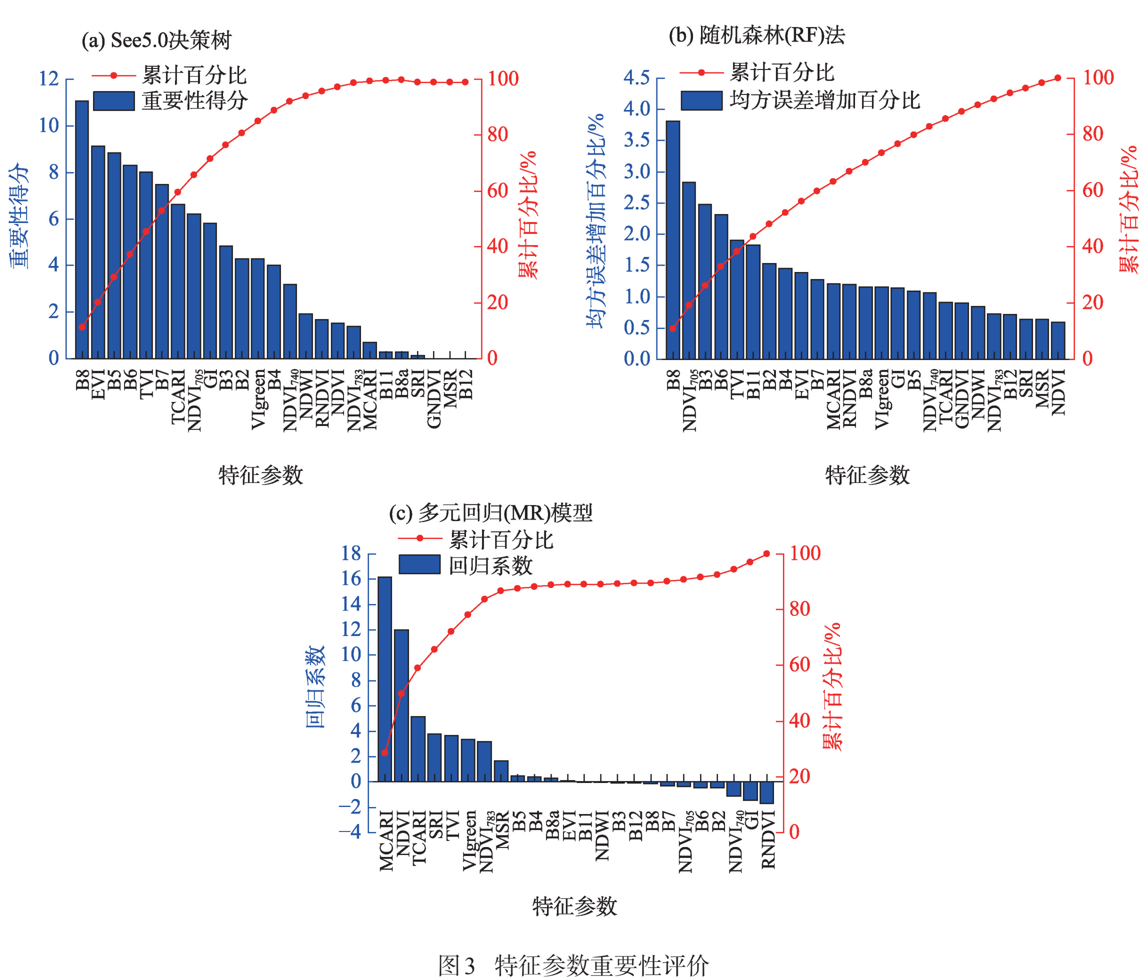


图3　特征参数重要性评价

Fig. 3　Importance evaluation of feature parameters

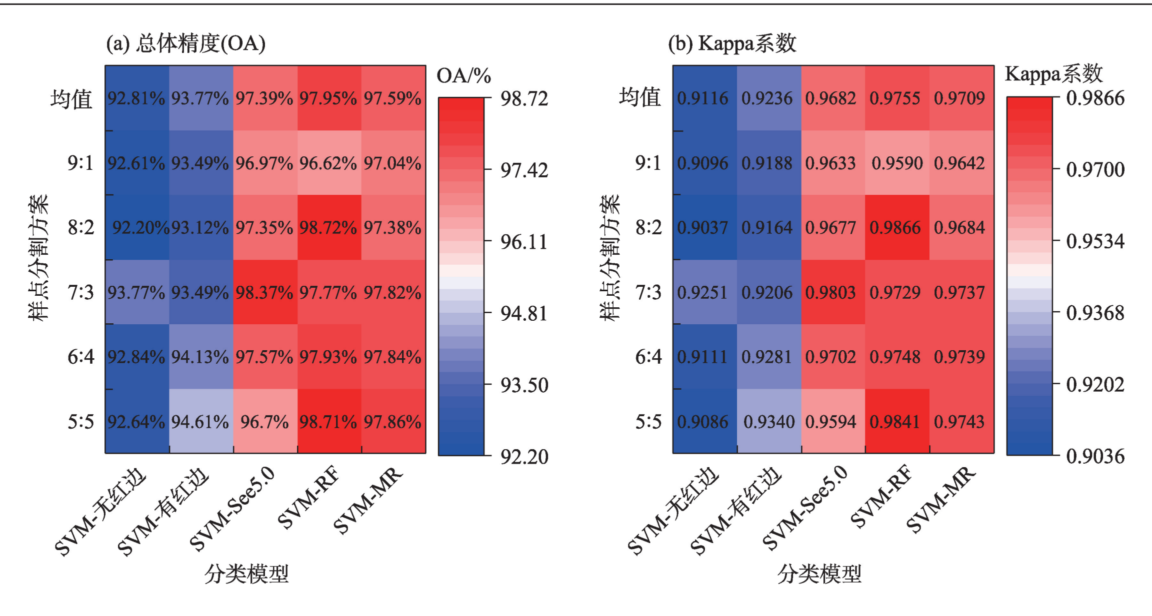


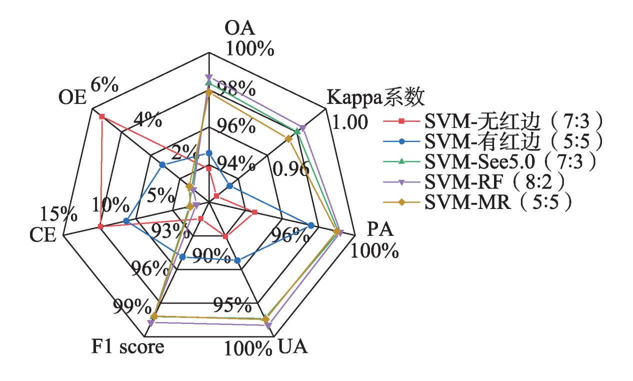
图4 不同分类模型总体精度和Kappa系数的热图

Fig. 4 Heat map of overall accuracy and Kappa coefficient for different classification models

中使用SVM算法开展干旱区农作物遥感分类研究是可行的。(2) 对比SVM-无红边和SVM-有红边的结果发现，加入红边波段后，农作物的OA和Kappa系数均有所上升，SVM-有红边分类模型的各样方分割方案OA均大于93.00%，Kappa系数均大于0.9100，由此可见在农作物遥感分类时加入红边波段可以提高分类精度。(3) 相较于SVM-有红边，SVM-RF、SVM-MR和SVM-See5.0方法在加入植被指数后，OA均值提高了4.18%、3.82%和3.62%，Kappa系数提高了0.0519、0.0473和0.0446。(4) 对比不同分类模型的均值发现，不同分类模型OA和Kappa的大小关系为：SVM-RF>SVM-MR>SVM-See5.0>SVM-有红边>SVM-无红边，说明SVM-RF分类模型的分类效果最好，相较于其他分类模型更适合在焉耆盆地开展农作物分类研究。

2.4 分类结果对比

选取每种分类模型与样方分割方案的最佳组合进行对比分析(图5)。由图5可知，5种最佳分类组合的OA关系为：SVM-RF(8:2)>SVM-See5.0(7:3)>SVM-MR(5:5)>SVM-有红边(5:5)>SVM-无红边(7:3)，5种最佳分类组合中分类效果最好的是SVM-RF分类模型在样方分割为8:2时取得的分类结果，其OA和Kappa系数最高，分别是98.72%和0.9866，表明SVM-RF(8:2)分类组合与地面实际情况具有较好的一致性；对比5种最佳分类组合的PA和UA精度发现，SVM-RF(8:2)的PA和UA最高，分别比SVM-MR



注：PA为制图精度；UA为用户精度；F1 score为F1分数；OE为漏分误差；CE为错分误差。

图5 5种最佳分类模型评价指标的雷达图

Fig. 5 Radar plot of evaluation indicators for the 5 best classification models

(5:5)、SVM-See5.0(7:3)、SVM-有红边(5:5)和SVM-无红边(7:3)的PA和UA高出0.21%和0.68%、0.09%和0.76%、1.60%和7.21%、4.68%和9.88%，表明SVM-MR(5:5)分类图中的结果与地面真实情况较为相符；F1 score的关系为SVM-RF(8:2)>SVM-See5.0(7:3)>SVM-MR(5:5)>SVM-有红边(5:5)>SVM-无红边(7:3)，说明SVM-RF(8:2)分类模型精确度较其他分类模型高。OE和CE的关系均为SVM-无红边(7:3)>SVM-有红边(5:5)>SVM-MR(5:5)>SVM-See5.0(7:3)>SVM-RF(8:2)，表明SVM-无红边(7:3)的“错分”和“漏分”现象较多，SVM-RF(8:2)的“错分”和“漏分”现象较少。综上，SVM-RF(8:2)分类模型的

分类结果在OA、Kappa系数、PA、UA、F1 score、CE和OE等方面表现最好，其分类效果优于其他分类模型。

研究借助PIE平台对5种最佳分类模型的分类结果进行制图，研究区作物空间分布如图6所示。对比图6c~e可见，3种分类模型的分类结果与当地实际数据吻合度较高。焉耆盆地农作物以辣椒、玉米、小麦和葡萄为主，种植面积分别约为1166.98 km^2、430.32 km^2、344.23 km^{2}和205.79 km^2，分别约占研究区耕地面积的43.58%、16.07%、12.85%和7.68%。整体上，焉耆盆地工业辣椒的种植面积最广，主要分布在焉耆回族自治县和博湖县；玉米、小麦和葡萄主要种植在和硕县；甜菜主要种植在和静县，水稻和果林主要种植在博湖县；兵团地区农作物种植分布较为集中，主要种植经济作物(工业辣椒和工业番茄)。综上，焉耆盆地主要农作物集中分布在土壤肥沃且灌溉设施较为完善的地区，种植面积较大且在研究区内连续分布，其他农作物种植较少且呈镶嵌式分布在研究区。

3　讨 论

本文借助PIE-Engine Studio平台，使用支持向量机算法，调用Sentinel-2时序数据，构建农作物时序植被指数，分割训练样本和检验样本，实现了对焉耆盆地农作物种植信息的提取与评估。PIE平台免费向公众开放，可以实时调用平台中存储的Sentinel-2遥感影像，在强大的云计算服务支持下，降低了遥感数据获取的门槛和计算成本，可满足不同尺度的研究需求[22]。研究结果在一定程度上解决了因干旱区生态环境脆弱、种植结构复杂且地块破

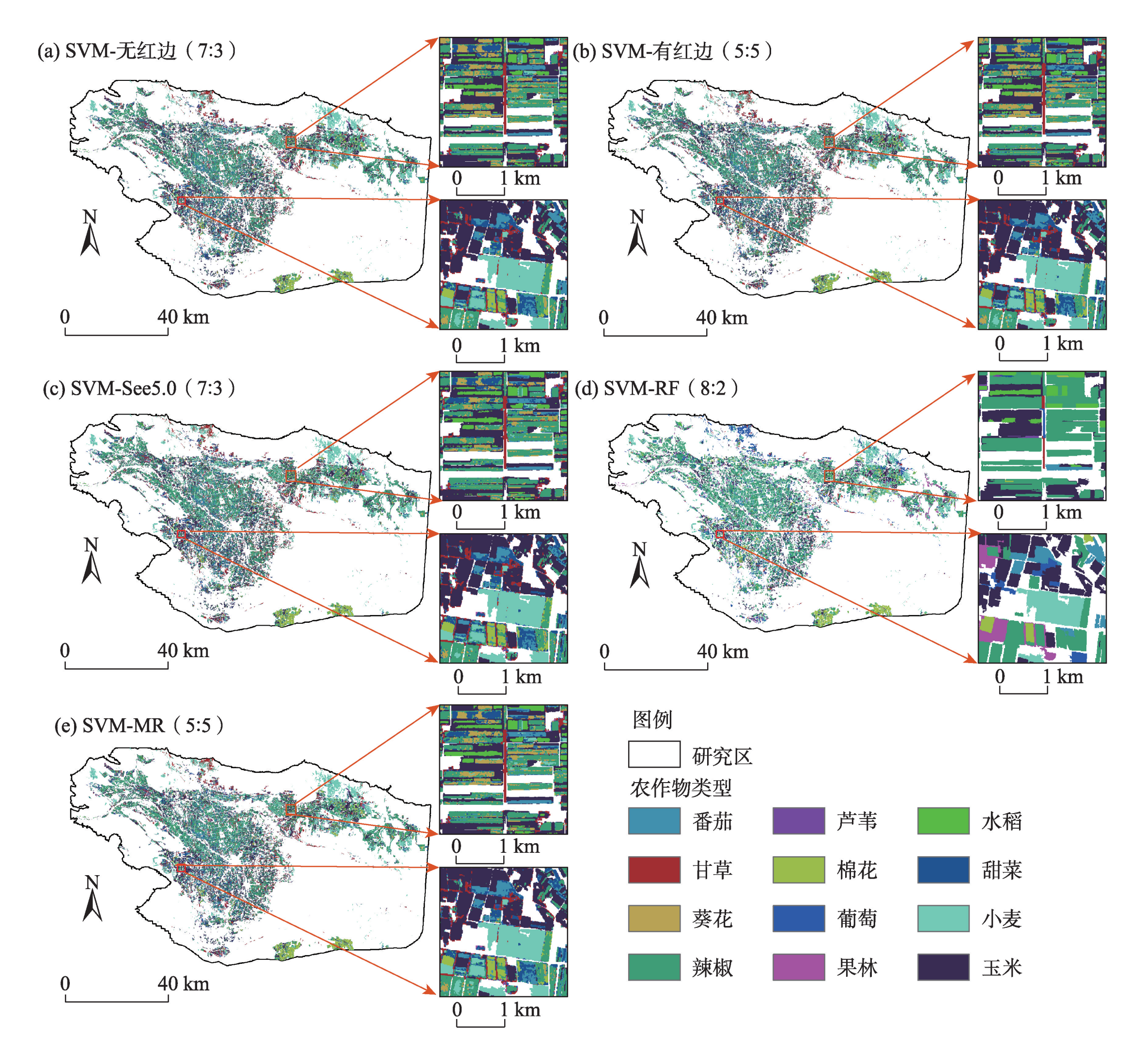


图6　5种分类模型的农作物分类结果

Fig. 6　Crop classification results for the five classification models

碎、农作物生长周期相近及光谱差异小等原因产生的分类精度较低的问题,为利用国产云平台实现干旱区农作物种植信息的快速、精确提取提供了新的思路和参考依据。

3.1 引入红边特征对分类精度的影响

从特征重要性结果可以看出,排名前十的特征参数中均有红边特征参数,表明红边特征参数在农作物分类识别领域具有明显优势,对于农作物生长状态和健康状况的判断具有重要的指导作用。研究中红边特征参数的引入使得SVM-有红边分类模型的OA和Kappa系数均值分别比SVM-无红边提高了0.96%和0.012。因此,在特征重要性分析中,排名前十的特征参数中出现红边特征参数是合理的,并且可以说明红边特征参数在农作物分类识别中的重要性。黄双燕等[28]和牛乾坤等[29]研究表明,在农作物遥感分类中引入红边光谱和红边指数,使机器学习分类器的OA得到提高,并使春、冬小麦的识别效果得到明显提升,表明红边光谱特征对分类精度起着决定性作用,河套灌区平均OA达到81%,Kappa系数达到0.68。综上,红边特征参数可以帮助分类器更灵敏地捕捉到不同农作物特有的生长特征和物候学差异从而提高分类模型的精度。

3.2 引入植被指数对分类精度的影响

研究中农作物植被指数的引入使得SVM-See5.0、SVM-RF和SVM-MR分类模型的OA均值比SVM-有红边分别提高了3.62%、4.18%和3.82%,Kappa系数均值分别提高了0.0446、0.0519和0.0473,说明植被指数在农作物分类中具有有效性和重要性,能够捕捉到农作物的生长特征并帮助区分不同农作物的类别。姜伊兰等[21]研究指出利用NDVI时序差异指数提取的农作物OA和Kappa系数比最大似然法和支持向量机法分别提高10.02%、0.21和4.18%和0.09;谷祥辉等[30]研究表明时间序列组合植被指数用于农作物分类是可行的,分类模型的精度和Kappa系数分别达到88.52%和0.8650;综上,多种植被指数结合时间序列信息可以从生长时期和生长状态上区分农作物,从而提高农作物识别的准确性。

研究使用了3种特征优选方法获得特征参数的重要性排序,排名前十的特征参数均含有红边波段或红边指数,表明红边波段和红边指数在农作物分类中发挥着重要作用。因此,在后续研究中将考虑更多与红边波段有关的植被指数,选取其中重要性较高的植被指数,提高作物分类精度。研究也存有不足,作物分类结果中部分农田存在“错分”和“漏分”现象,后续需通过人工干预和增加分类特征来减少“错分”和“漏分”现象的发生;另外,本次研究影像数据较为单一,后续考虑将Sentinel-2影像与其他遥感影像进行融合[31-32],充分挖掘地物的光谱特征、纹理特征和物候特征,进一步提高农作物识别精度。

4 结 论

(1) 使用PIE平台可快速地访问海量遥感图像及其他数据资源,因其高性能的云计算能力可快速完成覆盖焉耆盆地影像数据的去云、裁剪及植被指数构建等处理。基于多时间序列影像源与PIE相结合可方便快捷地进行作物种植信息提取,较本地处理具有明显优势。

(2) 25种分类模型的OA和Kappa系数均在92.20%和0.9037以上,其中SVM-RF方法在样方分割为8:2时,OA和Kappa系数最高,分别为98.72%和0.9866,说明在PIE中使用SVM算法提取作物信息是可行的。

(3) 5种SVM算法分类组合的OA均值关系为SVM-RF(97.95%)>SVM-MR(97.59%)>SVM-See5.0(97.39%)>SVM-有红边(93.77%)>SVM-无红边(92.81%),Kappa系数均值关系为SVM-RF(0.9755)>SVM-MR(0.9709)>SVM-See5.0(0.9682)>SVM-有红边(0.9236)>SVM-无红边(0.9116),其中SVM-RF、SVM-See5.0和SVM-MR方法中加入了红边波段和植被指数,提高了作物识别的精度。

(4) 焉耆盆地农作物以辣椒、玉米、小麦和葡萄为主,种植面积分别约为1166.98 km^2、430.32 km^2、344.23 km^2和205.79 km^2,分别约占研究区耕地面积的43.58%、16.07%、12.85%和7.68%,集中分布在土壤肥沃且灌溉设施较为完善的地区,种植面积较大且在研究区内连续分布,其他农作物种植较少,呈现镶嵌式分布。

参考文献(References)

[1] 唐华俊, 吴文斌, 杨鹏, 等. 农作物空间格局遥感监测研究进展[J]. 中国农业科学, 2010, 43(14): 2879-2888. [Tang Huajun, Wu

Wenbin, Yang Peng, et al. Recent progresses in monitoring crop spatial patterns by using remote sensing technologies[J]. Scientia Agricultura Sinica, 2010, 43(14): 2879-2888.]

[2] 陈亚宁, 李玉朋, 李稚, 等. 全球气候变化对干旱区影响分析[J]. 地球科学进展, 2022, 37(2): 111-119. [Chen Yaning, Li Yupeng, Li Zhi, et al. Analysis of the impact of global climate change on dryland areas[J]. Advances in Earth Science, 2022, 37(2): 111-119.]

[3] 刘珍环, 杨鹏, 吴文斌, 等. 近30年中国农作物种植结构时空变化分析[J]. 地理学报, 2016, 71(5): 840-851. [Liu Zhenhuan, Yang Peng, Wu Wenbin, et al. Spatio-temporal changes in Chinese crop patterns over the past three decades[J]. Acta Geographica Sinica, 2016, 71(5): 840-851.]

[4] 胡琼, 吴文斌, 宋茜, 等. 农作物种植结构遥感提取研究进展[J]. 中国农业科学, 2015, 48(10): 1900-1914. [Hu Qiong, Wu Wenbin, Song Qian, et al. Recent progresses in research of crop patterns mapping by using remote sensing[J]. Scientia Agricultura Sinica, 2015, 48(10): 1900-1914.]

[5] 杨雪峰. 使用高分遥感影像获取塔里木河胡杨高度信息[J]. 遥感技术与应用, 2021, 36(5): 1199-1208. [Yang Xuefeng. Estimation height of *Populus euphratica* in Tarim River using VHR satellite images[J]. Remote Sensing Technology and Application, 2021, 36(5): 1199-1208.]

[6] 梁继, 郑镇炜, 夏诗婷, 等. 高分六号红边特征的农作物识别与评估[J]. 遥感学报, 2020, 24(10): 1168-1179. [Liang Ji, Zheng Zhenwei, Xia Shiting, et al. Crop recognition and evaluation using red edge features of GF-6 satellite[J]. Journal of Remote Sensing, 2020, 24(10): 1168-1179.]

[7] Jia K, Wu B F, Li Q Z. Crop classification using HJ satellite multispectral data in the North China Plain[J]. Journal of Applied Remote Sensing, 2013, 7(1): 73576, doi: 10.1117/1.JRS.7.073576.

[8] 美合日阿依·莫一丁, 买买提·沙吾提, 李金朝. 基于Sentinel-2时间序列数据及物候特征的棉花种植区提取[J]. 干旱区地理, 2022, 45(6): 1847-1859. [Moyiding Meiheriayi, Shawuti Maimaiti, Li Jinzhao. Extraction of cotton planting area based on Sentinel-2 time series data and phenological characteristics[J]. Arid Land Geography, 2022, 45(6): 1847-1859.]

[9] Liangzhi Y, Stanley W, Ulrike W S. Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach[J]. Agricultural Systems, 2008, 99(2): 126-140.

[10] Chad M, Navin R, Jonathan A F. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000[J]. Global Biogeochemical Cycles, 2008, 22(1): 1-19.

[11] Reichstein M, Camps V G, Stevens B, et al. Deep learning and process understanding for data-driven Earth system science[J]. Nature, 2019, 566(7743): 195-204.

[12] 许晴, 张锦水, 张凤, 等. 深度学习农作物分类的弱样本适用性[J]. 遥感学报, 2022, 26(7): 1395-1409. [Xu Qing, Zhang Jinshui, Zhang Feng, et al. Applicability of weak samples to deep learning crop classification[J]. Journal of Remote Sensing, 2022, 26(7): 1395-1409.]

[13] 贾银江, 姜涛, 苏中滨, 等. 基于改进SVM算法的典型作物分类方法研究[J]. 东北农业大学学报, 2020, 51(7): 77-85. [Jia Yinjiang, Jiang Tao, Su Zhongbin, et al. Study on classification method of typical crops based on improved SVM algorithm[J]. Journal of Northeast Agricultural University, 2020, 51(7): 77-85.]

[14] 梁习卉子, 陈兵旗, 李民赞, 等. 基于HOG特征和SVM的棉花行数动态计数方法[J]. 农业工程学报, 2020, 36(15): 173-181. [Liang Xihuizi, Chen Bingqi, Li Minzan, et al. Method for dynamic counting of cotton rows based on HOG feature and SVM[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(15): 173-181.]

[15] 边增淦, 王文, 江渊. 黑河流域中游地区作物种植结构的遥感提取[J]. 地球信息科学学报, 2019, 21(10): 1629-1641. [Bian Zenggan, Wang Wen, Jiang Yuan. Remote sensing of cropping structure in the middle reaches of the Heihe River Basin[J]. Journal of Geoinformation Science, 2019, 21(10): 1629-1641.]

[16] 田鑫, 何海, 金双彦, 等. 基于遥感影像的张掖灌区作物种植结构提取研究[J]. 中国农村水利水电, 2022(8): 206-212, 217. [Tian Xin, He Hai, Jin Shuangyan, et al. Crop planting structure extraction in Zhangye irrigation area based on remote sensing images[J]. China Rural Water and Hydropower, 2022(8): 206-212, 217.]

[17] 郭其乐, 李军玲, 郭鹏. 基于作物双时相遥感特征的花生种植区提取[J]. 应用气象学报, 2022, 33(2): 218-230. [Guo Qile, Li Junling, Guo Peng. Extraction of peanut planting area based on dual-temporal remote sensing features of crops[J]. Journal of Applied Meteorological Science, 2022, 33(2): 218-230.]

[18] 付东杰, 肖寒, 苏奋振, 等. 遥感云计算平台发展及地球科学应用[J]. 遥感学报, 2021, 25(1): 220-230. [Fu Dongjie, Xiao Han, Su Fenzhen, et al. Remote sensing cloud computing platform development and Earth science application[J]. Journal of Remote Sensing, 2021, 25(1): 220-230.]

[19] 刘通, 任鸿瑞. GEE平台下利用物候特征进行面向对象的水稻种植分布提取[J]. 农业工程学报, 2022, 38(12): 189-196. [Liu Tong, Ren Hongrui. Object-oriented extraction of paddy rice planting areas using phenological features from the GEE platform[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(12): 189-196.]

[20] 潘力, 夏浩铭, 王瑞萌, 等. 基于Google Earth Engine的淮河流域越冬作物种植面积制图[J]. 农业工程学报, 2021, 37(18): 211-218. [Pan Li, Xia Haoming, Wang Ruimeng, et al. Mapping of the winter crop planting areas in Huaihe River Basin based on Google Earth Engine[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(18): 211-218.]

[21] 姜伊兰, 陈保旺, 黄玉芳, 等. 基于Google Earth Engine和NDVI时序差异指数的作物种植区提取[J]. 地球信息科学学报, 2021, 23(5): 938-947. [Jiang Yilan, Chen Baowang, Huang Yu-

fang, et al. Crop planting area extraction based on Google Earth Engine and NDVI time series difference index[J]. Journal of Geoinformation Science, 2021, 23(5): 938−947.]

[22] 程伟, 钱晓明, 李世卫, 等. 时空遥感云计算平台 PIE-Engine Studio的研究与应用[J]. 遥感学报, 2022, 26(2): 335−347. [Cheng Wei, Qian Xiaoming, Li Shiwei, et al. Research and application of PIE-Engine Studio for spatiotemporal remote sensing cloud computing platform[J]. Journal of Remote Sensing, 2022, 26(2): 335−347.]

[23] 田颖, 陈卓奇, 惠凤鸣, 等. 欧空局哨兵卫星Sentinel−2A/B数据特征及应用前景分析[J]. 北京师范大学学报(自然科学版), 2019, 55(1): 57−65. [Tian Ying, Chen Zhuoqi, Hui Fengming, et al. ESA Sentinel−2A/B satellite: Characteristics and applications [J]. Journal of Beijing Normal University (Natural Science Edition), 2019, 55(1): 57−65.]

[24] 齐红超, 祁元, 徐瑱. 基于C5.0决策树算法的西北干旱区土地覆盖分类研究——以甘肃省武威市为例[J]. 遥感技术与应用, 2009, 24(5): 648− 653, 553. [Qi Hongchao, Qi Yuan, Xu Zhen. The study of the northwest arid zone land-cover classification cased on C5.0 decision tree algorithm at Wuwei City, Gansu Province[J]. Remote Sensing Technology and Application, 2009, 24(5): 648−653, 553.]

[25] 许晴, 张锦水, 张凤, 等. 深度学习农作物分类的弱样本适用性[J]. 遥感学报, 2022, 26(7): 1395−1409. [Xu Qing, Zhang Jinshui, Zhang Feng, et al. Applicability of weak samples to deep learning crop classification[J]. Journal of Remote Sensing, 2022, 26(7): 1395−1409.]

[26] 张善红, 白红英, 齐贵增, 等. 基于多元线性回归模型和Anusplin的秦巴山区≥10 ℃积温空间模拟比较[J]. 水土保持研究, 2022, 29(1): 184−189, 196. [Zhang Shanhong, Bai Hongying, Qi Guizeng, et al. Spatial simulation of active accumulated temperature ≥10 ℃ in Qinling-Daba Mountains based on Anusplin and multiple linear regression model[J]. Research of Soil and Water Conservation, 2022, 29(1): 184−189, 196.]

[27] Giorgos M, Jungho I, Caesar O. Support vector machines in remote sensing: A review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 66(3): 247−259.

[28] 黄双燕, 杨辽, 陈曦, 等. 机器学习法的干旱区典型农作物分类[J]. 光谱学与光谱分析, 2018, 38(10): 3169−3176. [Huang Shuangyan, Yang Liao, Chen Xi, et al. Study of typical arid crops classification based on machine learning[J]. Spectroscopy and Spectral Analysis, 2018, 38(10): 3169−3176.]

[29] 牛乾坤, 刘浏, 黄冠华, 等. 基于GEE和机器学习的河套灌区复杂种植结构识别[J]. 农业工程学报, 2022, 38(6): 165−174. [Niu Qiankun, Liu Liu, Huang Guanhua, et al. Extraction of complex crop structure in the Hetao Irrigation District of Inner Mongolia using GEE and machine learning[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(6): 165−174.]

[30] 谷祥辉, 张英, 桑会勇, 等. 基于哨兵2时间序列组合植被指数的作物分类研究[J]. 遥感技术与应用, 2020, 35(3): 702−711. [Gu Xianghui, Zhang Ying, Sang Huiyong, et al. Research on crop classification method based on Sentinel−2 time series combined vegetation index[J]. Remote Sensing Technology and Application, 2020, 35(3): 702−711.]

[31] 郭交, 朱琳, 靳标. 基于Sentinel−1和Sentinel−2数据融合的农作物分类[J]. 农业机械学报, 2018, 49(4): 192−198. [Guo Jiao, Zhu Lin, Jin Biao. Crop classification based on data fusion of Sentinel−1 and Sentinel−2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(4): 192−198.]

[32] 薛朝辉, 钱思羽. 融合Landsat 8与Sentinel−2数据的红树林物候信息提取与分类[J]. 遥感学报, 2022, 26(6): 1121−1142. [Xue Zhaohui, Qian Siyu. Fusion of Landsat 8 and Sentinel−2 data for mangrove phenology information extraction and classification[J]. Journal of Remote Sensing, 2022, 26(6): 1121−1142.]

Remote sensing identification and assessment of crops in the Yanqi Basin, Xinjiang, China based on Sentinel−2 time series data

ZHANG Xuhui[1], Yusufujiang RUSULI[1,2], QIU Zhongli[1],
Yaxiaer AISIKEER[1], Abudureheman WUSIMAN[1],
(1. Watershed Information Integration and Ecological Security Laboratory, College of Geographic Science and Tourism, Xinjiang Normal University, Urumqi 830054, Xinjiang, China;
2. Key Laboratory of Lake Environment and Resources in the Arid Zone of Xinjiang, Urumqi 830054, Xinjiang, China)

Abstract: To obtain timely and accurate information about crop cultivation in arid zones, this study used the PIE-Engine Studio platform to extract 14 vegetation indices in the Yanqi Basin, Xinjiang, China based on 2022 Sentinel−2 images and 1948 field location sampling data during the crop reproduction period. Crop planting information was extracted using the See5.0 decision tree, random forest (RF), and multiple regression (MR) models to select feature parameters. Each model was combined with support vector machine (SVM) algorithms to construct five classification models and five sample segmentation schemes. The best classification scheme was determined by visual interpretation and confusion matrix comparison. The results are as follows: (1) The overall accuracy (OA) and Kappa coefficients of all classification models are above 92.20% and 0.9037, respectively, indicating that it is feasible to extract crop information using the SVM algorithm in the PIE platform. (2) The mean OA and Kappa coefficients of SVM-with-red-edge are 93.77% and 0.9236, which are 0.96% and 0.0120, respectively. (3) The introduction of vegetation index improved the OA and Kappa coefficients of SVM-RF, SVM-MR, and SVM-See5.0 compared with the SVM-with-red-edge method. (4) The mean OA and Kappa coefficient relationships for the five classification models were SVM-RF>SVM-MR>SVM-See5.0>SVM-with-red-edge>SVM-without-red-edge, showing that the inclusion of the red-edge band and vegetation index significantly improved the accuracy of crop identification, with SVM-RF (8:2) being the best classification model with OA and Kappa coefficients of 98.72% and 0.9866, respectively. These results provide new ideas and references for accurate and rapid access to large-scale arid zone crop information.

Key words: crop; Sentinel−2; support vector machines; PIE-Engine Studio; Yanqi Basin